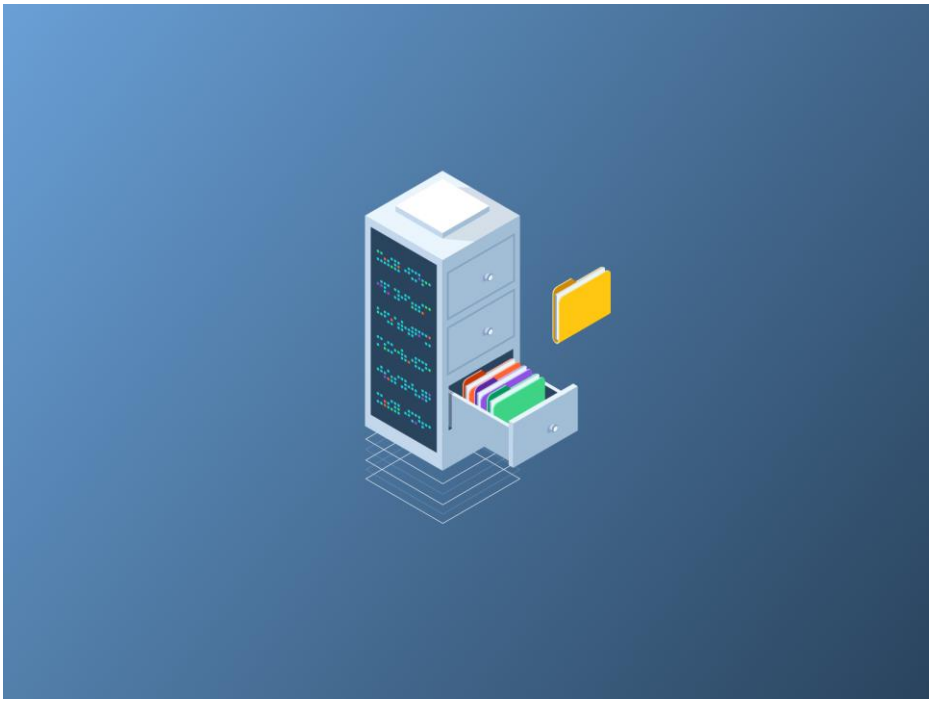


ASSIGNMENT 1

ITRI 613

ABSTRACT

This assignment is about different database management systems and what sets them apart, as well as their part in the industry.

Enrico Dreyer

31210783

Table of Contents

PART 1

First three database management systems:

	MSSQL Server	MySQL	Maria DB
Data Structure	T-SQL	SQL	SQL
Licensing	Proprietary	GNU Generally Public License	GNU Generally Public License
Scalability	Vertical complex	Vertical complex	Vertical
Main Advantages	Variety of versions. End-to-end business data solution. Cloud database support.	Free installation. Simple syntax and mild complexity. Cloud-compatible.	Encryption. Broad functionality. High performance.
Main Disadvantages	Cost-consuming, Unclear and floating license conditions, Complicated tuning process	Scalability challenges. Partial open source. Limited compliance with SQL standards.	Still growing community. Gaps between MySQL and MariaDB update versions.
Relational /Extended-relational	Relational database management system.	Relational database management system.	Relational database management system.

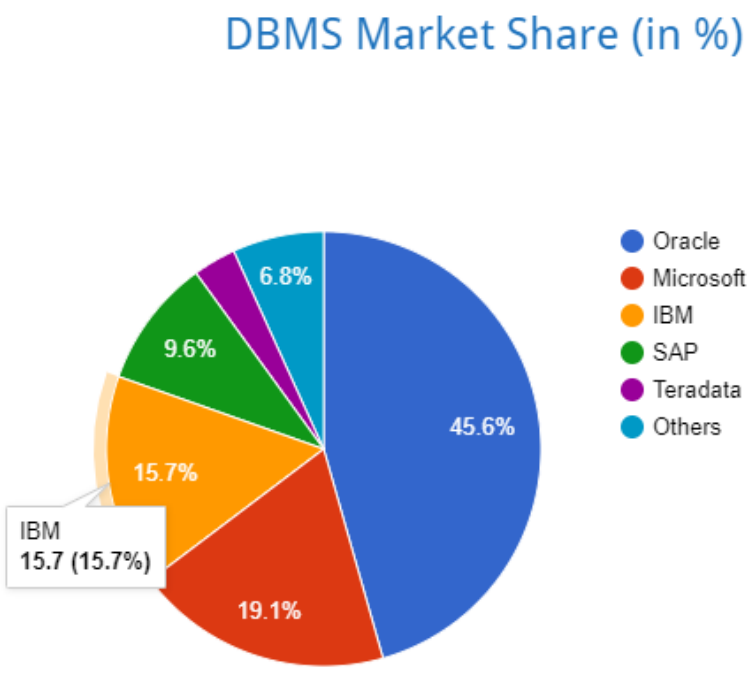
Next four database management systems:

	Oracle	Cassandra	Redis	MongoDB
Data Structure	Multi-model, SQL	NoSQL, column orientated.	NoSQL, key-value	NoSQL, document orientated.
Licensing	Proprietary	Open source	Open source, BSD 3-clause	SSPL
Scalability	Vertical	Horizontal	Horizontal	Horizontal
Main Advantages	Innovations for daily workflow. Strong tech support and documentation Large capacity.	Data security. Flexibility and on-hand amendments.	Rapid solution. Massive data processing.	Simple data access, storage, input, and retrieval. Easy compatibility with other data models. Horizontally scalable solution.

Main Disadvantages	High cost, Resource-consuming. Hard learning curve.	Slow reading. Requires additional resources.	Requires dataset to fit into memory. No support for query language or joins.	Extensive memory consumption. Data insecurity. Complicated process to interpret into other query languages.
Relational /Extended-relational	Relational database management system.	Extended-Relational database management system. (NoSQL)	Extended-Relational database management system. (NoSQL)	Extended-Relational database management system. (NoSQL)

Part 2

DBMS Market



1. Amazon

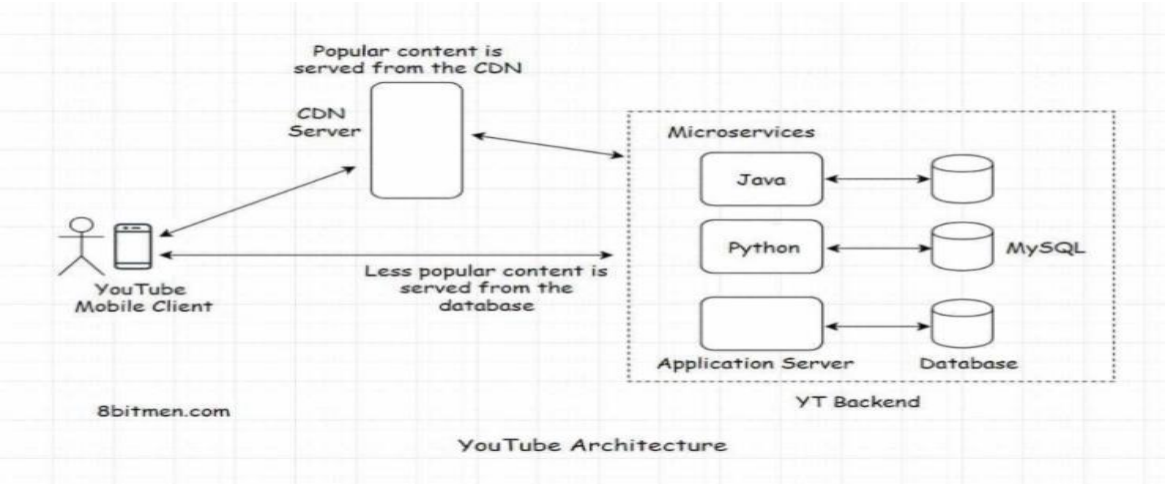
In 2007 Amazon announced their own NoSql databases called SimpleDB (Walters, 2019). Although Amazon also uses DynamoDb, Redshift, Aurora and some legacy Oracle for their own retail operations (not the ones that AWS provides for its own customers).

Amazon uses the “right tool for the job”, they use the right DBMS depending on the job. The teams use the one that makes more sense to them.

Amazon does not use a Relational database management system (RDBMS) to store data of their products, because it simply does not support the required scale.

2. YouTube

Youtube’s backend is written in C, C++, Java and Python. MySQL is the primary DBMS, powered by Vitess (Shivang, 2019). For caching they use Memcache and for node co-ordination they use Zookeeper.



The popular videos are served from the CDN server as seen in the image above. The videos that are played less are fetched from the database.

YouTube started as a single database instance. As the queries per second increased, the developers had to scale the relational database horizontally.

3. Google

Google uses spanner and it is their own globally distributed RDBMS, the successor to BigTable (Chang, 2008). Although google claims that is not a pure relational system because the tables still require a primary key.

Google uses BigTable to store data, they use it for projects such as Google Earth, web indexing and Google Finance.

They use this because it works across many machines, it is fast and is a large-scale DBMS.

The tables consist of columns and rows, and every cell has a time stamp. Each table is a multidimensional sparse map. Because BigTable does not support rich SQL queries nor supports joins it makes BigTable a non-relational database.

4. Facebook

Facebook uses a number of database management systems, including MySQL, Haystack, Cassandra and HBase (What Database does Facebook use 2019, 2019).

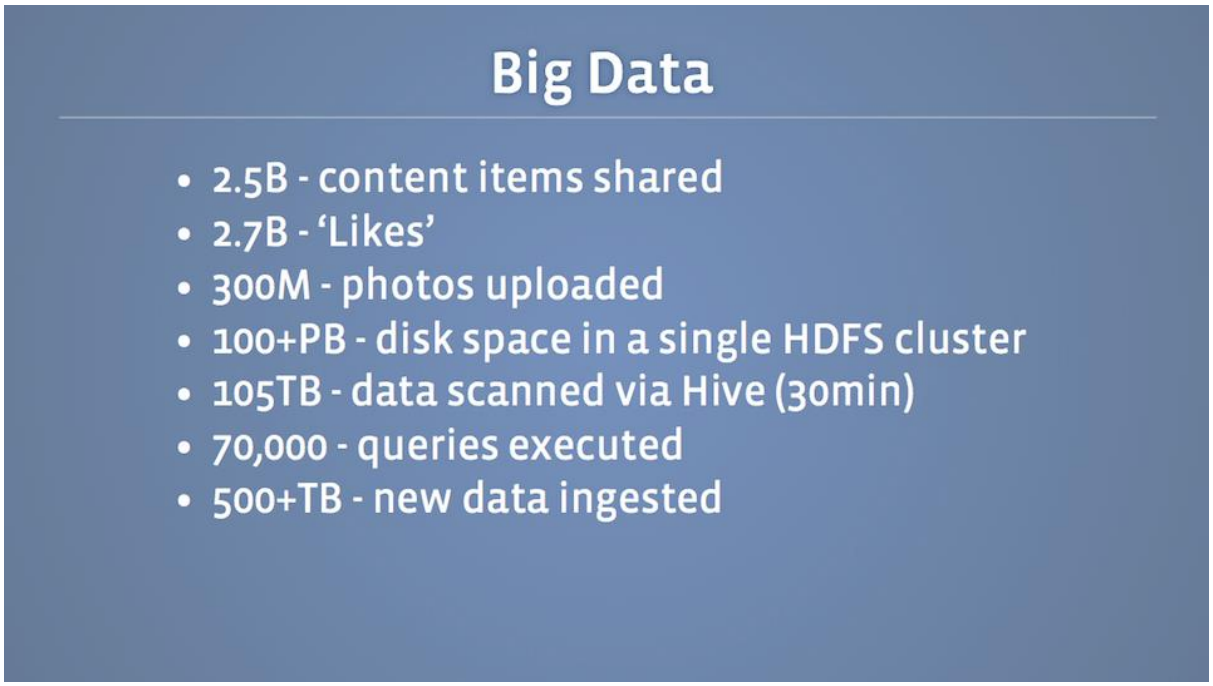
MySql is used as a primary DBMS, to store all structured data, such as wall posts, timelines, and information of various users.

HBase is used for their Facebook messenger. The data and information are physically fragmented and are then termed as regions.

Cassandra is the method that they use for the purpose of managing big amounts of data structure that expands to different commodity servers. For example, they use this for inbox search.

Haystack is used to save pictures, for every picture that is uploaded they save four images in different sizes. Haystack photo infrastructure has the ability to combine photo serving tier and storage tier together to create a new tier. A HTTP base server is used to store pictures in what is called a universal object store.

Memcached is the process used by Facebook for in-memory solution that is used only of caching.



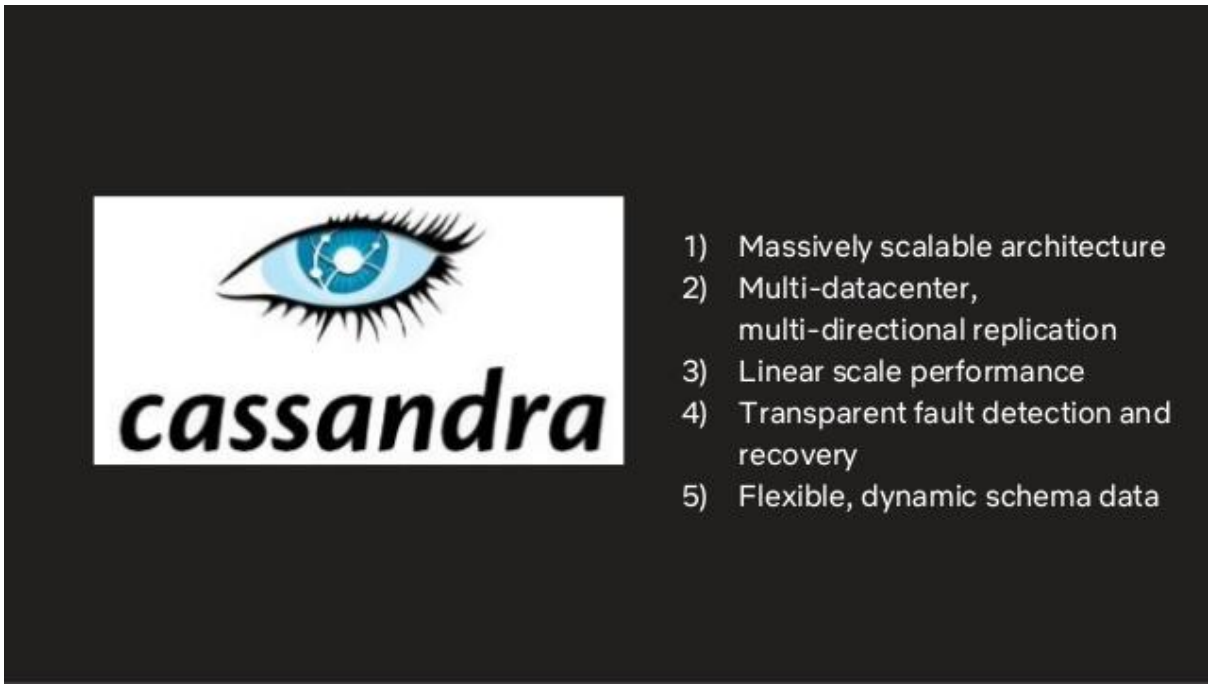
5. Netflix

Netflix uses SimpleDB, Casandra and HBase (Finley, 2011). These are all NoSQL DBMS.

They use SimpleDB because it is highly durable and was an obvious choice because they use AWS cloud. This also allows them to use handy data format and query features, for example multiple attributes per row key, consistent reads and batch operations.

HBase is used because it is deeply integrated with Hadoop. This allows the ability to combine batch map-reduce Hadoop jobs with real-time HBase queries, by making use of HDFS as a shared storage platform.

Cassandra is used for its lack of single points of failure, its scalability, and cross-regional deployments. They use this because they can scale dynamically and horizontally by adding more servers, without having to re-shard or reboot, unlike a distributed database.



References

Chang, F. (2008, June). *Bigtable: A Distributed Storage System for Structured Data*. Retrieved from ACM Transactions on Computer Systems (TOCS): https://dl.acm.org/doi/abs/10.1145/1365815.1365816

Finley, K. (2011, January 28). *Why and How Netflix Adopted NoSQL Databases*. Retrieved from readwrite: https://readwrite.com/2011/01/28/how-netflix-adopted-nosql/?__cf_chl_jschl_tk__=ccd31774b18ec56eddbe70d34eb26816e4b9a868-1616358007-0-ARX44y6aOWxP8sqIwY320AvJwgA1zc2kJEbyyH3AEu5VGh080bGrvl7Y9da8Nwl7MZlBxgkGScA2bgS7ptM9__z1kQ8VQKy4OSzqsT1LSdyDdlZHEHN9zPzeM

Shivang. (2019, May). *YouTube Database – How Does It Store So Many Videos Without Running Out Of Storage Space?* Retrieved from REAL LIFE ARCHITECTURES: https://www.8bitmen.com/youtube-database-how-does-it-store-so-many-videos-without-running-out-of-storage-space/#:~:text=YouTube%E2%80%99s%20backend%20microservices%20are%20written%20using%20Python,%20C,,used%20for%20caching%20&%20Zookeeper%20for%20node%20

Walters, M. (2019). *What database does Amazon use to store product information?* Retrieved from Quora: https://www.quora.com/What-database-does-Amazon-use-to-store-product-information

What Database does Facebook use 2019. (2019, May 21). Retrieved from Gangboard: https://www.gangboard.com/blog/what-database-does-facebook-use/#:~:text=%20What%20Database%20does%20Facebook%20use%202019?%20,the%20messages%20back%20in%20the%20previous...%20More